



More efficient tests robust to heteroskedasticity of unknown form

Emmanuel Flachaire

► To cite this version:

Emmanuel Flachaire. More efficient tests robust to heteroskedasticity of unknown form. *Econometric Reviews*, 2005, 24 (2), pp.219-241. 10.1081/ETC-200067942 . halshs-00175914

HAL Id: halshs-00175914

<https://shs.hal.science/halshs-00175914>

Submitted on 1 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

More Efficient Tests Robust to Heteroskedasticity of Unknown Form

by

Emmanuel Flachaire

EUREQUA, Université Paris I Panthéon-Sorbonne

March 2005

Abstract

In the presence of heteroskedasticity of unknown form, the Ordinary Least Squares parameter estimator becomes inefficient and its covariance matrix estimator inconsistent. Eicker (1963) and White (1980) were the first to propose a robust consistent covariance matrix estimator, that permits asymptotically correct inference. This estimator is widely used in practice. Cragg (1983) proposed a more efficient estimator, but concluded that tests based on it are unreliable. Thus, this last estimator has not been used in practice. This paper is concerned with finite sample properties of tests robust to heteroskedasticity of unknown form. Our results suggest that reliable and more efficient tests can be obtained with the Cragg estimators in small samples.

JEL classification : C12, C52

Key-words : wild bootstrap, heteroskedasticity-robust test, regression model

Corresponding author: Emmanuel Flachaire, EUREQUA - Maison des Sciences Economiques, 106-112 bd de l'Hopital, 75647 Paris Cedex 13, France. e-mail: Emmanuel.Flachaire@univ-paris1.fr

I would like to thank two anonymous referees for comments that improved substantially the paper.

1 Introduction

Let us consider the linear heteroskedastic model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (1)$$

where \mathbf{y} is a n -vector containing the values of the dependent variable, \mathbf{X} a $n \times k$ matrix in which each column is an explanatory variable, assumed to be exogenous, and $\boldsymbol{\beta}$ is a k -vector of parameters. With heteroskedastic error terms, its covariance matrix $\boldsymbol{\Omega}$ is diagonal, with typical t^{th} diagonal element σ_t^2 . In the presence of heteroskedasticity of unknown form, the matrix $\boldsymbol{\Omega}$ cannot be estimated consistently and the Generalized Least Squares (GLS) estimator, which is an efficient estimator, cannot be computed. The Ordinary Least Squares (OLS) parameter estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad (2)$$

is inefficient but still consistent, and its OLS covariance matrix estimator is inconsistent in general. Thus, conventional t and F tests do not have their namesake distributions, even asymptotically, under the null hypothesis that they test.

Eicker (1963) and White (1980) proposed a *Heteroskedasticity Consistent Covariance Matrix Estimator* (HCCME), which permits asymptotically correct inference on $\boldsymbol{\beta}$ in the presence of heteroskedasticity of unknown form. MacKinnon and White (1985) considered a number of possible forms of the estimator. They showed that the HCCME's, as well as the associated t or F statistics, can be seriously biased in finite sample, especially in the presence of observations with high leverage. See also Chesher and Jewitt (1987), who show that the extent of the bias is related to the structure of the regressors. However, reliable inferences can be obtained in large samples, and this estimator is widely used in practice.

Cragg (1983) proposed a more efficient estimator than the OLS parameter estimator, based on over-identification restrictions. This estimator can be viewed as a Generalized Method of Moments (GMM) estimator with instruments taken as: regressors, their cross-products and successive positive powers. Furthermore, the Eicker-White estimator is a particular case of the Cragg estimator, with instruments taken as regressors and with no additional instruments. In his paper, Cragg showed that a gain in efficiency can be obtained using few additional instruments. However, a major problem is encountered: a downward bias of the covariance matrix estimator appears to be serious and disappears very slowly as the number of observations increases. In addition, tests based on this estimator have poor finite sample behavior: large errors in the rejection probability are exhibited. Consequently, this estimator is rarely used in practice.

To date, published studies have focused on the reliability of tests based on the Eicker-White estimator. To achieve reliability in small samples, two approaches have been proposed: computing the covariance estimator using restricted residuals or/and using bootstrap methods, see Davidson and MacKinnon (1985), Horowitz (1997) and Godfrey and Orme (2004). In this paper, we study both the reliability and the efficiency of tests based on Eicker-White and on Cragg estimators. First, our simulation results suggest that reliable inferences based on Cragg estimators can be obtained in small samples when we combine the two

previous approaches. In addition, our results suggest that by using restricted residuals rather than unrestricted residuals, the power increases substantially. Moreover, we review the Cragg estimator and propose a more judicious choice of instruments. Indeed, the efficient GLS estimator is equivalent to the Cragg estimator using the instruments $\Omega^{-1}\mathbf{X}$. However, we cannot use these instruments, because the matrix Ω is unknown and cannot be estimated consistently. We propose to use instruments which include the components of the matrix $\Omega^{-1}\mathbf{X}$: this leads us to use regressors, their *inverses*, their cross-products and *cross-divisions*, their successive positive and *negative* powers as instruments.

In section 2, we present the Eicker-White and Cragg Estimators as GMM estimators, and we discuss the choice of instruments. In section 3, we describe our experimental design in detail. Section 4 presents our simulation results and section 5 is concerned with the number of instruments to be used in practice. Finally, Section 6 concludes.

2 Heteroskedasticity of unknown form

In this section, we present the Eicker-White and Cragg estimators that are robust to heteroskedasticity of unknown form, and we make some proposals on the choice of instruments.

2.1 The Eicker-White and Cragg robust estimators

The Eicker-White and Cragg estimators can be encompassed by the Generalized Method of Moments (GMM) approach. Let \mathbf{W} be an $n \times l$ matrix of valid instruments, with a typical row \mathbf{W}_t , such that $E(u_t | \mathbf{W}_t) = 0$. A necessary condition for the efficiency of the Generalized Moments Estimators, showed by Hansen (1982), is that it minimizes the following criterion function :

$$(\mathbf{y} - \mathbf{X}\beta)^\top \mathbf{W}(\mathbf{W}^\top \Omega \mathbf{W})^{-1} \mathbf{W}^\top (\mathbf{y} - \mathbf{X}\beta)$$

The first-order conditions for a minimum of this function are

$$\mathbf{X}^\top \mathbf{W}(\mathbf{W}^\top \Omega \mathbf{W})^{-1} \mathbf{W}^\top (\mathbf{y} - \mathbf{X}\beta) = 0$$

which leads to the GMM estimator

$$\hat{\beta}_{\text{GMM}} = (\mathbf{X}^\top \mathbf{W}(\mathbf{W}^\top \Omega \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}(\mathbf{W}^\top \Omega \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{y} \quad (3)$$

Under regularity conditions, $n^{1/2}(\hat{\beta}_{\text{GMM}} - \beta_0)$ is asymptotically normally distributed with mean vector 0 and covariance matrix

$$\text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^\top \mathbf{W}(\mathbf{W}^\top \Omega \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X} \right)^{-1} \quad (4)$$

Hansen (1985) argued that the optimal instrument choice is $\mathbf{W} = \Omega^{-1}\mathbf{X}$: with that choice, the GMM estimator is asymptotically efficient, and reduces to the Generalized

Least Squares (GLS) estimator. A second choice may be $\hat{\Omega}$, when $\hat{\Omega}$ is a consistent estimator of Ω . Any other choice of instruments leads to an inefficient estimator. In the case of heteroskedasticity of unknown form, the diagonal matrix Ω cannot be estimated consistently. However, operational versions of the estimators (3) and (4) have been proposed, based on consistent estimators of $n^{-1}\mathbf{W}^\top\Omega\mathbf{W}$.

Eicker (1963) and White (1980) proposed a *Heteroskedasticity Consistent Covariance Matrix Estimator*, which is equal to the GMM estimator (4) with the regressors taken as instruments $\mathbf{W} = \mathbf{X}$. In that case, the parameter estimate (3) reduces to the OLS parameter estimate,

$$\hat{\beta}_{\text{White}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad (5)$$

and its covariance matrix (4) can be estimated by

$$\widehat{\text{Var}}(\hat{\beta}_{\text{White}}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \quad (6)$$

where the $n \times n$ diagonal matrix $\hat{\Omega}$ has a typical diagonal element $a_t^2 \hat{u}_t^2$, and \hat{u}_t is the t^{th} OLS residual. MacKinnon and White (1985) considered a number of possible forms of this estimator:

$$HC_0 : a_t = 1 \quad HC_2 : a_t = (1 - h_t)^{-1/2} \quad (7)$$

$$HC_1 : a_t = (n/(n - k))^{1/2} \quad HC_3 : a_t = (1 - h_t)^{-1} \quad (8)$$

where $h_t \equiv \mathbf{X}_t(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_t^\top$, the t^{th} element of the orthogonal projection matrix on to the span of the columns of \mathbf{X} . The basic version proposed by Eicker and White corresponds to HC_0 . The HC_1 version corresponds to the use in the homoskedastic case of the unbiased OLS estimator of the error variance. The HC_2 version corresponds to the use in the homoskedastic case of unbiased OLS estimator of the variance of \hat{u}_t , which is proportional to $1 - h_t$. Finally, arguments based on the jackknife lead MacKinnon and White to propose form HC_3 . Simulation results by MacKinnon and White (1985) show that, in terms of error of the rejection probability (ERP), or size distortion, HC_0 is outperformed by HC_1 , which is in turn outperformed by HC_2 and HC_3 . The HC_3 -based estimator is widely used in practice.

Cragg (1983) proposed using a *feasible* version of the GMM estimator (3), where the regressors \mathbf{X} , their cross-products and successive positive powers, are taken as instruments \mathbf{W} , and where the diagonal matrix Ω is replaced by $\hat{\Omega}$,

$$\hat{\beta}_{\text{Cragg}} = (\mathbf{X}^\top \mathbf{W} (\mathbf{W}^\top \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} (\mathbf{W}^\top \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{y}, \quad (9)$$

its covariance matrix (4) can be estimated by

$$\widehat{\text{Var}}(\hat{\beta}_{\text{Cragg}}) = \left(\mathbf{X}^\top \mathbf{W} (\mathbf{W}^\top \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X} \right)^{-1} \quad (10)$$

It is clear that the Eicker-White estimator is a particular case of the Cragg estimator: if $\mathbf{W} = \mathbf{X}$, equations (9) and (10) are respectively equal to equations (5) and (6). Cragg showed that a gain in efficiency can be obtained when additional instruments are taken

in \mathbf{W} . However, large size distortions of tests based on Cragg estimators are exhibited in finite samples. Because of its poor finite sample behavior, this approach has not been used in practice.

Implementation of this estimator can be made by an instrumental variables (IV) estimation, where \mathbf{y}^* is the dependent variable, \mathbf{X}^* the matrix of regressors and \mathbf{W}^* the matrix of instruments,

$$\mathbf{y}^* \equiv U^{-1}\mathbf{y} \quad \mathbf{X}^* \equiv U^{-1}\mathbf{X} \quad \mathbf{W}^* \equiv U\mathbf{W} \quad (11)$$

and $U = \text{diag}(a_1\hat{u}_1, \dots, a_n\hat{u}_n)$. The IV parameter estimate is equal to (9). The IV covariance matrix estimate should be multiplied by the inverse of the variance of the error terms estimate $1/\hat{\sigma}_{iv}^2$ in order to be equal to (10). Computation of this estimator can be made in two steps. First, we estimate the model (1) by OLS, and obtain the residuals \hat{u}_t . Then, we estimate the model described above (11) by IV. Note that this implementation requires all residuals to be different from zero, otherwise (9) and (10) should be computed directly.

2.2 Choice of instruments

We have seen above that if we could estimate the matrix $\mathbf{\Omega}$ consistently, the optimal choice of instruments is $\mathbf{\Omega}^{-1}\mathbf{X}$: with that choice, the GMM estimator reduces to the efficient GLS estimator. If we cannot estimate the matrix $\mathbf{\Omega}$ consistently, we cannot use these optimal instruments, but we can judiciously choose instruments such that the space spanned by them $\mathcal{S}(\mathbf{W})$ is as close as possible as the space spanned by the optimal instruments $\mathcal{S}(\mathbf{\Omega}^{-1}\mathbf{X})$. This suggests using the components of $\mathbf{\Omega}^{-1}\mathbf{X}$ as instruments and leads us to use the instruments proposed by Cragg, and in addition, inverses, cross-divisions and successive negative powers of regressors.

Proposition 1 *Let us consider the regression model (1). In the presence of heteroskedasticity of unknown form, a more efficient estimator than the Ordinary Least Squares parameter estimate is the parameter estimate (9) with the following instruments: regressors, their inverses, cross-products, cross-divisions, successive positive and negative powers. For all values of regressors equal to zero, the corresponding values of instruments are equal to zero.*

The following discussion can be helpful to better understand why we should get better results by including inverses, cross-divisions and successive negative powers. Let \mathbf{Q} be the regressors' matrix of the skedastic function $\sigma_t^2 = h(\mathbf{Q}_t)$. From Weierstrass' Theorem: any continuous function on a closed interval is the uniform limit of a polynomial series. In other words, it means that we can approximate any continuous function by a polynomial with a sufficiently high degree. This suggests approximating the skedastic function by a linear combination of the variables \mathbf{Q}_t , cross-products and higher powers. Heteroskedasticity is a particular problem for usual OLS inference when it is related to the regressors \mathbf{X}_t , as when $\mathbf{Q}_t = \mathbf{X}_t$. Cragg (1983) suggests using regressors, their cross-products and higher powers as instruments.

Rather than considering the elements of the skedastic function, we consider the elements of the matrix $\mathbf{\Omega}^{-1}\mathbf{X}$. For example, let us consider $\mathbf{Q}_t = [\mathbf{x}_{t1}, \mathbf{x}_{t2}]$ where \mathbf{x}_{t1} and \mathbf{x}_{t2} are two elements of two n -vectors of the regressors. We thus obtain the polynomial approximation:

$$\sigma_t^2 \equiv \alpha_0 + \alpha_1 \mathbf{x}_{t1} + \alpha_2 \mathbf{x}_{t2} + \alpha_3 \mathbf{x}_{t1} \mathbf{x}_{t2} + \alpha_4 \mathbf{x}_{t1}^2 + \alpha_5 \mathbf{x}_{t2}^2 \quad (12)$$

where α_i , for $i = 0, \dots, 5$, are unknown parameters. The element of the t^{th} row and j^{th} column in the matrix $\mathbf{\Omega}^{-1}\mathbf{X}$ is $\{\mathbf{\Omega}^{-1}\mathbf{X}\}_{tj} = \mathbf{x}_{tj}/\sigma_t^2$. We have to consider two cases. In the first case, if $\mathbf{x}_{tj} = 0$ then $\{\mathbf{\Omega}^{-1}\mathbf{X}\}_{tj} = 0$. In the second case, if $\mathbf{x}_{tj} \neq 0$ then $\{\mathbf{\Omega}^{-1}\mathbf{X}\}_{tj}$ can be approximated by

$$\frac{\mathbf{x}_{tj}}{\sigma_t^2} \equiv \left(\alpha_0 \frac{1}{\mathbf{x}_{tj}} + \alpha_1 \frac{\mathbf{x}_{t1}}{\mathbf{x}_{tj}} + \alpha_2 \frac{\mathbf{x}_{t2}}{\mathbf{x}_{tj}} + \alpha_3 \frac{\mathbf{x}_{t1}\mathbf{x}_{t2}}{\mathbf{x}_{tj}} + \alpha_4 \frac{\mathbf{x}_{t1}^2}{\mathbf{x}_{tj}} + \alpha_5 \frac{\mathbf{x}_{t2}^2}{\mathbf{x}_{tj}} \right)^{-1} \quad (13)$$

A polynomial approximation of this function would be a linear combination of its variables, their cross-products and higher powers. The space spanned by these variables taken as instruments $\mathcal{S}(\mathbf{W})$ is an approximation of the space spanned by $\mathcal{S}(\mathbf{\Omega}^{-1}\mathbf{X})$. It follows that a judicious choice of instruments is regressors, their inverses, cross-products, cross-divisions and successive positive and negative powers. For all values of regressors equal to zero, corresponding observations of instruments are equal to zero.

3 Experimental design

In this section, we present our model design, that is, our choices of regression models, regressors and instruments. We then explain how we compute the power in Monte Carlo experiments. Finally, we present an appropriate bootstrap method for the case of heteroskedasticity of unknown form.

3.1 Model design

In our experiments, we consider the linear regression model

$$\mathbf{y}_t = \beta_0 + \beta_1 \mathbf{x}_{1t} + \beta_2 \mathbf{x}_{2t} + u_t, \quad t = 1, \dots, n, \quad (14)$$

in which the explanatory variables are supposed to be fixed in repeated samples. The error terms are assumed to be mutually independent and to have a common mean of zero, but they may be heteroskedastic, with $E(u_t^2) = \sigma_t^2$. We wish to test the null hypothesis that $\beta_1 = 0$ with a t -statistic

$$\tau = \frac{\hat{\beta}_1 - 0}{\left[\widehat{\text{Var}}(\hat{\beta}_1) \right]^{1/2}} \quad (15)$$

based on the Eicker-White estimators (5) and (6), or the Cragg estimators (9) and (10).

Since (14) is a linear model and $\boldsymbol{\tau}$ is independent of the choices of β_0 and β_2 , we set $\beta_0 = \beta_2 = 0$ without a loss of generality. We can rewrite $u_t = \sigma_t v_t$, where $E(v_t^2) = 1$. Thus, under the null hypothesis, our data are generated by a DGP of the form

$$\mathbf{y}_t^* = \sigma_t v_t, \quad t = 1, \dots, n. \quad (16)$$

Many different choices for the heteroskedasticity can be used, but a function of regressors should be included. We set $\sigma_t^2 = \mathbf{x}_{1t}^2$ because this choice gives rise to strong heteroskedasticity, which leads to a serious bias of the OLS covariance matrix (White 1980). The v_t are independent mean zero variables of unit variance, and will be normal in the experiments. Our experiments are performed with a sample of size $n = 100$; the number of Monte Carlo replications is equal to $N = 10,000$.

- *Choice of regressors:*

Chesher and Jewitt (1987) showed that finite sample performance is very sensitive to the structure of regressors: if the diagonal elements of the projection matrix on the space spanned by the regressors (h_t) are not well balanced, the Eicker-White estimators perform badly. MacKinnon and White (1985) studied the finite sample performance of tests based on the Eicker-White estimators, using Monte Carlo simulations. They used 50 real economic observations as regressors, and repeated those 50 observations when the sample size increased. They found good results as n increases: it comes from the fact that by repeating the data the influence of high leverage observations decreases quickly as the number of observations increases. In a similar study, Cragg (1983) generated regressors drawn from a lognormal distribution and found very slow improvement as n increases. This comes from the fact that the lognormal distribution is very skewed and some regressors took large values compared to the others, even in a large sample. It follows that the influence of high leverage observations decreases slowly as the number of observations increases.

These results highlight a major point: the choice of the regressors is more important than the choice of the sample size. Indeed, we could find very good results from a small sample with well balanced regressors and very bad results from a large sample with unbalanced regressors. This leads us to consider different regressors' schemes in our experiments, with more or less well-balanced data. We draw a pseudo-random variable η_t from the standard normal distribution and use the following transformation to generate a regressor,

$$\eta_t \sim N(0, 1) \quad \text{and} \quad \mathbf{x}_t = \boldsymbol{\kappa}^{\eta_t} \quad (17)$$

It is clear that, if $\boldsymbol{\kappa} = \exp(1)$, regressors are drawn from a lognormal distribution, and they are not well-balanced. On the other hand, if $\boldsymbol{\kappa} \approx 1$, the regressors are well-balanced. In our experiments, we use different values of $\boldsymbol{\kappa}$:

$$\boldsymbol{\kappa} = 1.1, 1.2, \dots, 2.5, 2.6, \exp(1), 2.8, 2.9, \dots, 3.5 \quad (18)$$

From Chesher and Jewitt's results, we expect that the simulation results will exhibit poor finite sample performance as $\boldsymbol{\kappa}$ increases, at least when using the Eicker-White estimator.

- *Choice of instruments:*

The total number of instruments l , varies across experiments. The $n \times l$ matrix of instruments \mathbf{W} includes regressors and additional instruments, successively appended in two

ways. The first choice of regressors is, as proposed by Cragg (1983),

$$\mathbf{W}_1^+ = [\boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_1^2, \mathbf{x}_2^2] \quad (19)$$

$$\mathbf{W}_2^+ = [\boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_1\mathbf{x}_2] \quad (20)$$

$$\mathbf{W}_3^+ = [\boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_1\mathbf{x}_2, \mathbf{x}_1^3, \mathbf{x}_2^3] \quad (21)$$

We successively introduce: 1- squared, 2- cross-products, 3- power three of regressors. The second choice of instruments is, as described in proposition 1,

$$\mathbf{W}_1^\pm = [\boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2, 1/\mathbf{x}_1, 1/\mathbf{x}_2] \quad (22)$$

$$\mathbf{W}_2^\pm = [\boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2, 1/\mathbf{x}_1, 1/\mathbf{x}_2, \mathbf{x}_1/\mathbf{x}_2, \mathbf{x}_2/\mathbf{x}_1, 1/(\mathbf{x}_1\mathbf{x}_2)] \quad (23)$$

$$\mathbf{W}_3^\pm = [\boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2, 1/\mathbf{x}_1, 1/\mathbf{x}_2, \mathbf{x}_1/\mathbf{x}_2, \mathbf{x}_2/\mathbf{x}_1, 1/(\mathbf{x}_1\mathbf{x}_2), \mathbf{x}_1\mathbf{x}_2, \mathbf{x}_1^2, \mathbf{x}_2^2] \quad (24)$$

We successively introduce: 1- division, 2- cross-divisions, 3- cross-products and squared of regressors.

3.2 Power calculation

In our model, we test the null hypothesis $\beta_1 = 0$. Then, we study the power of a test statistic by drawing sample data from a DGP under the alternative hypothesis $\beta_1 \neq 0$,

$$\mathbf{y}_t^\diamond = \beta_1 \mathbf{x}_{1t} + \sigma_t v_t, \quad t = 1, \dots, n \quad (25)$$

We test the null hypothesis $\beta_1 = 0$, *not true*, with the test statistic (15). It is clear that the power depends on the exact value of β_1 .

It is interesting to note that, in our experimental design, it is not necessary to generate a sample under the alternative hypothesis. Let us assume that $\hat{\beta}_1^*$ is the OLS parameter estimate computed from a sample simulated under the null DGP (16), $(\mathbf{y}^*, \boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2)$, and $\hat{\beta}_1^\diamond$ is the OLS parameter estimate computed from a sample simulated under the alternative DGP (25), $(\mathbf{y}^\diamond, \boldsymbol{\iota}, \mathbf{x}_1, \mathbf{x}_2)$, with the same error terms v_t , for $t = 1, \dots, n$. Then, we can show

$$\tau^\diamond = \frac{\hat{\beta}_1^\diamond - 0}{[\widehat{\text{Var}}(\hat{\beta}_1^\diamond)]^{1/2}} = \frac{\hat{\beta}_1^* + \beta_1}{[\widehat{\text{Var}}(\hat{\beta}_1^*)]^{1/2}} \quad (26)$$

This result comes from the fact that $\hat{\beta}_1^\diamond = \hat{\beta}_1^* + \beta_1$ and $\widehat{\text{Var}}(\hat{\beta}_1^\diamond) = \widehat{\text{Var}}(\hat{\beta}_1^*)$, which can be easily shown. It allows us to compute different values of τ^\diamond with different values of β_1 , without generating samples under the alternative hypothesis.

To compare the power of alternative test statistics, it is not desirable to plot power against the nominal level if all the tests do not have the correct size and if they exhibit significant ERP. It is thus desirable to plot power against the “true” level. In our experiments, we compute the power at a same rejection probability level; this is sometimes called size-corrected tests.

3.3 Bootstrap method

The dismal performance of asymptotic tests based on these estimators, see Cragg (1983), suggests the use of better test procedures such as bootstrap methods. An appropriate version of the bootstrap for the case of heteroskedasticity of unknown form is the wild bootstrap, proposed by Liu (1988), following the suggestion of Wu (1986) and Beran (1986). For a detailed discussion on this method, see Horowitz (2000) and Davidson and Flachaire (2001). Bootstrap methods normally rely on simulations based on an artificial data-generating process, constructed as close as possible to the true DGP under the null hypothesis. The wild bootstrap data-generating process, or bootstrap DGP, is of the form

$$\mathbf{y}_t^* = \tilde{\beta}_0 + \tilde{\beta}_2 \mathbf{x}_{2t} + a_t \ddot{u}_t \varepsilon_t^* \quad (27)$$

where $\tilde{\beta}_0$ and $\tilde{\beta}_2$ are the OLS parameter estimates under the model (14) that respects the null, \ddot{u}_t is the OLS restricted or unrestricted residual and a_t takes the forms HC_0 , HC_1 , HC_2 or HC_3 , see (8). We can set $\tilde{\beta}_0 = \tilde{\beta}_2 = 0$, because the test statistic is independent of the choice of these two parameters. Simulation results by Davidson and Flachaire (2001) suggest that it is a mistake to mix unrestricted residuals in the estimators and restricted residuals in the bootstrap DGP. Thus, we choose to take the same function $a_t \ddot{u}_t$ in the Eicker-White and Cragg estimators and in the wild bootstrap DGP. The ε_t^* are mutually independent drawings, completely independent of the original data, from some auxiliary distribution, such that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$. In our experiments, we use the Rademacher distribution, as recommended by Davidson and Flachaire (2001)

$$\mathbf{F} : \quad \varepsilon_t^* = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2, \end{cases} \quad (28)$$

The simulation results of these latter authors, with those of Godfrey and Orme (2001) show that this two-point distribution outperforms the other distributions proposed in the literature. In addition, simulation results in Flachaire (2005) suggest that this version of the wild bootstrap outperforms different versions of another appropriate bootstrap method - the pairs bootstrap (Freedman 1981, Flachaire 1999) - and is the only one for which the bootstrap test always gives better results than the asymptotic test. We therefore use this version of the wild bootstrap in our experiments. A P -value bootstrap can be obtained by the following steps:

1. Estimate the model (14) under the alternative and under the null by Ordinary Least Squares (OLS): we obtain vectors of parameter estimates and residuals. Then, compute a realization $\hat{\tau}$ of the t statistic that tests the null, based on Cragg's estimators.
2. Generate a bootstrap sample, based on the bootstrap data-generating process (27), where ε_t^* is drawn from the two-point distribution \mathbf{F} . Compute Cragg's estimators, and then, a realization $\hat{\tau}_j^*$ of the t statistic that tests the null. Repeat this step B times.

The proportion of the B realizations $\hat{\tau}_j^*$, $j = 1 \dots B$, greater than $\hat{\tau}$ is a P -value bootstrap. In our experiments, we use $B = 499$; see Davidson and MacKinnon (2000) for a detailed discussion on the choice of the number of bootstrap replications.

4 Simulation results

In this section, we present the Monte Carlo results on the Error of the Rejection Probability, or size distortion, and on the Power of the tests based on the Eicker-White and Cragg estimators. First, we find similar results to those of Cragg, with a similar model design. We propose successive improvements to get the size correct, and then we go on to the question of power.

• Cragg results

At first, we use a similar model design as in Cragg (1983). Then, we consider a test statistic, as defined in (15), based on the Eicker-White and Cragg estimators (see equations 5, 6, 9, 10), with no transformation of unrestricted residuals HC_0 (see 7) and with different choices of instruments:

- 1– **Cragg1** corresponds to the use of W_1^+ , see (19)
- 2– **Cragg2** corresponds to the use of W_2^+ , see (20)
- 3– **Cragg3** corresponds to the use of W_3^+ , see (21)

Figure 1 shows the Error in the Rejection Probability (*ERP*), or size distortion, at the nominal level $\alpha = 0.05$, of asymptotic tests based on the Eicker-White and Cragg estimators, as a function of κ , a measure of the presence of high leverage observations in the regressors, defined in (17). It is clear from that figure that all tests exhibit very large ERP as κ increases. For instance, **Cragg3** shows an *ERP* close to 0.25 when the regressors are drawn from a lognormal distribution ($\kappa = \exp(1)$): this means that, if the null is true, the test is supposed to reject the null at a nominal level equals to 5%, but in fact the test rejects the null at a “true” level close to 30%. In other words, inference is not at all reliable. In addition, we can see from this figure that the Eicker-White test (**White**) exhibits smaller ERP than the Cragg tests using additional instruments, and the ERPs strongly increase as the number of instruments increases. Recall that the Eicker-White test is a particular case of the Cragg test, with regressors taken as instruments and with no additional instruments. Finally, these results are similar to those of Cragg (1983): they suggest that tests based on Cragg estimators with extra instruments are unreliable in finite samples.

• Restricted residuals

An initial approach to controlling finite sample significance levels of heteroskedastic-robust tests has been proposed by Davidson and MacKinnon (1985). These authors propose to compute the test statistic by replacing the unrestricted residuals by the corresponding restricted residuals. Their simulation results suggest that it gives more reliable inference. Figure 2 shows ERPs, at the nominal level $\alpha = 0.05$, of the Eicker-White and Cragg tests based on the *restricted* residuals, as a function of κ . We can see from this figure a substantial reduction in the ERPs compared to the use of the unrestricted residuals (figure 1). The ERP of **Cragg3** is very small with or without well balanced regressors, that is for all values of κ . As in Davidson and MacKinnon’s results (1985), we can see that the Eicker-White test based on restricted residuals slightly under-rejects the null (the ERPs are slightly negative in figure 2), and the Eicker-White test based on unrestricted residuals largely over-rejects

(the ERPs are positive in figure 1). These results suggest that the use of restricted residuals substantially decreases ERPs of tests based on Cragg estimators. However, as noted by Godfrey and Orme (2004), a major drawback of using restricted residuals is that each null hypothesis requires a specific estimation of the Cragg estimator. For instance, testing the significance of each regressor requires k different Cragg covariance matrix estimates.

• Bootstrap method and unrestricted residuals

A second approach to controlling finite sample significance levels of heteroskedastic-robust tests, which does not require different estimations of the covariance matrix for different null hypotheses has been investigated by Horowitz (1997), Davidson and Flachaire (2001) and Godfrey and Orme (2001). Their simulation results suggest that the ERP of t tests based on Eicker-White estimators can be substantially removed by using the wild bootstrap method. Figure 3 shows the ERPs, at a nominal level $\alpha = 0.05$, of the Eicker-White and Cragg bootstrap tests, as a function of κ . The heteroskedasticity-robust estimators are computed with unrestricted residuals, that is, as in the Cragg scheme. Compared to the corresponding asymptotic tests (figure 1), we can see from this figure that the ERPs of all the tests are substantially removed by using the wild bootstrap. These results suggests that ERPs of asymptotic tests based on Cragg estimators are substantially reduced by the use of the wild bootstrap.

• Bootstrap method and restricted residuals

In our preceding results, we have seen that ERPs of Cragg tests can be substantially reduced using two approaches: with restricted residuals or with the wild bootstrap. However, significant ERPs are still visible in some cases and in particular for high values of κ , that is, in the presence of high leverage observations. Godfrey and Orme (2004) have shown recently that neither of these two approaches is by itself sufficient in the cases of several linear restrictions on regression coefficients. These authors have shown that the two approaches must be combined to achieve reliability. Note that the major drawback of using restricted residuals would also appear to be present in this case. In addition, Davidson and Flachaire (2001) have shown that an extra term is introduced in their Edgeworth expansion of the ERP of the Eicker-White test computed with the wild bootstrap, if unrestricted residuals are used. Figure 4 shows the ERPs, at a nominal level $\alpha = 0.05$, of the Eicker-White and Cragg tests computed with the restricted residuals and with the wild bootstrap. We can see from this figure that ERPs of all the tests are very close to zero for all values of κ . As suggested by Godfrey and Orme (2004) and Davidson and Flachaire (2001), these results suggest that a combination of the two approaches provides reliable inference in finite sample, in the presence or not of high leverage observations. They suggest the remarkable conclusion that heteroskedastic-robust tests based on Cragg estimators with additional instruments can be used in practice, if they are computed with restricted residuals and with the wild bootstrap.

• Power

A test must be able to discriminate between the null and the alternative hypothesis. The ERP, or size distortion, is concerned with the rejection of the null, when the null is true;

the power is concerned with the rejection of the null, when the null is false. Therefore, once we obtain a test that gets the size correct, we can go on to the question of power. This can lead us to re-examine the question of computing Cragg estimators with unrestricted or restricted residuals. Indeed, we could argue that when the null is true, the use of restricted residuals should provide a better approximation of the distribution of the error terms. Thus, it would be better to use restricted residuals to compute the Cragg estimator. However, several authors, including van Giersbergen and Kiviet (2002), argue that when the null is false, the unrestricted residuals should provide a better approximation of the distribution of the error terms, and thus, should improve the power. This last argument is not supported in the simulation experiments, see MacKinnon (2002) and Flachaire (2004, 2005). Thus, a study on the use of restricted residuals compared to unrestricted residuals makes sense.

- *Unrestricted residuals:* figure 5 shows the power of Eicker-White and Cragg tests computed with unrestricted residuals and with the wild bootstrap, as a function of β_1 , at a rejection probability $RP = 0.05$ and with lognormal regressors $\kappa = \exp(1)$. The power increases as β_1 move away from 0, and if $\beta_1 = 0$, the power is equal to the rejection probability 0.05. Thus, the most powerful test would reject the null as soon as $\beta_1 \neq 0$ and would take the form \top in our figure, with the bottom of the vertical segment at 0.05 and the horizontal segment at 1.0. In addition, we plot the power of a test based on the Generalized Least Square estimator, because the GLS estimator is asymptotically efficient. This last method is preferred in the presence of heteroskedasticity of known form, we use it as a benchmark, for comparison purposes only. From this figure, we can see that the GLS test is the most powerful and that the Eicker-White test exhibits a large lack of power. Our experiments display the same results as Cragg (1983): we can see a significant increase in power with few additional instruments, but a decrease in power with too many instruments.

- *Restricted residuals:* figure 6 shows the power of Eicker-White and Cragg tests computed with restricted residuals and with the wild bootstrap. Compared to the use of unrestricted residuals (figure 5), all the tests show a large increase in power. We can see that the power of the Eicker-White test is clearly the worst and is outperformed by the Cragg tests. The power of the Cragg tests increases as the number of instruments increases.

We have seen previously that using restricted residuals provides a decrease of ERP. In addition, our results suggest that using restricted residuals provides a substantial increase in power. They suggest that Cragg estimators can be used in practice to achieve an efficiency gain over Eicker-White estimators.

• Transformed residuals

MacKinnon and White (1985) have proposed different versions of the Eicker-White estimator to improve the finite sample performance of tests based on it; see (7) and (8). Their simulation results, with those of Long and Ervin (2000), give strong evidence that HC_3 outperforms the other transformations when we use unrestricted residuals. Thus, we perform new experiments using the HC_3 version of the transformed residuals to compute the Cragg estimators with the wild bootstrap, but the results are not reported here. When we use unrestricted residuals, we find similar conclusions: HC_3 outperforms the other

transformations. However, when we use restricted residuals, there are no clear differences whether we use the transformation HC_3 of the residuals or not. However, in some cases this transformation gives slightly better results; thus, we use the HC_3 transformation in the following experiments.

• New instruments

The discussion on the choice of instruments in section 2.2 leads us to consider additional instruments which extend those proposed by Cragg (1983), as described in proposition 1. Thus, here we consider the Eicker-White and Cragg estimators, with the HC_3 transformation of the restricted residuals and with different choices of instruments:

- 1– **Cragg1n** corresponds to the use of W_1^\pm , see (22)
- 2– **Cragg2n** corresponds to the use of W_2^\pm , see (23)
- 3– **Cragg3n** corresponds to the use of W_3^\pm , see (24)

Simulation results show that Cragg tests here also have reasonable size properties, but the results are not reported. Figure 7 shows the power of Eicker-White and Cragg tests computed with the HC_3 transformation of the restricted residuals and with the wild bootstrap, at the rejection probability $RP = 0.05$ and with regressors drawn from a lognormal distribution $\kappa = \exp(1)$. We can see that a large gain of power is obtained by the use of these new instruments, compared to using the instruments proposed by Cragg (figure 6). It is interesting to see that **Cragg1n** in figure 7 is more powerful than **Cragg3** in figure 6. In other words, the Cragg tests based on the regressors and the addition of *only* the inverse of the regressors taken as instruments (W_1^\pm) exhibit more power than the Cragg tests with regressors and the addition of squares/cross-products and regressors to the power three taken as instruments (W_3^+). Furthermore, the addition of cross-divisions (**Cragg2n**) and cross-products/successive powers (**Cragg3n**) slightly increases the power. We can see that the power of the Cragg tests using these new instruments is close to the power of the GLS test.

These results suggest that the use of divisions/cross-divisions and successive negative powers of the regressors, in addition to the squares/cross-products and additional powers, substantially increases the power of the Cragg tests in such a way that the finite sample performance can become close to that of the GLS test.

• F -test

It is interesting to check if our results hold using a several linear restrictions on regression coefficients. In such cases, Godfrey and Orme (2004) have shown that using the restricted residuals and the wild bootstrap must be combined to achieve reliability. We consider a null hypothesis with more than one restriction, $H_0 : \beta_1 = \beta_2 = 0$, and F -tests based on Cragg estimators. Table 1 presents the ERP and POWER results based on lognormal regressors, $\kappa = \exp(1)$, and two different schemes:

Cragg scheme: F -test based on unrestricted residuals, HC_0 and W^+

New scheme: F -test based on restricted residuals, HC_3 , W^\pm and the wild bootstrap.

The left column, named ERP, in table 1, shows the ERP, or size distortion, of a F -test at

	ERP		Power	
	<i>Cragg scheme</i>	<i>New scheme</i>	<i>Cragg scheme</i>	<i>New scheme</i>
White	0.4073	-0.0003	0.4887	0.7503
Cragg1	0.5554	0.0050	0.2957	0.7181
Cragg2	0.5559	-0.0116	0.3001	0.9996
Cragg3	0.5934	-0.0002	0.3242	0.9984

Table 1: Empirical level and Power of the F -test using the White and Cragg estimators

a nominal level $\alpha = 0.05$. We can see that the F -tests based on the Cragg scheme exhibit very large ERPs (more than 0.4073), and that the F -tests based on the New scheme have very small ERPs (less than 0.0116). The right column, named **Power**, in table 1, shows the power of a F -test at a “true” level $RP = 0.05$, for fixed values of β_1 and β_2 under the alternative. We choose $\beta_1 = \beta_2 = 0.02$ in order to obtain distinct results between the different tests. The power of a test based on the GLS estimators is equal to 1. We can see that the F -tests based on the Cragg scheme have small power compared to that of a GLS test (far from 1) and that the F -tests based on the New scheme have a power close to a GLS test (close to 1) using additional instruments (Cragg2 Cragg3).

5 How many instruments?

Estimators with good finite sample properties would be as efficient as possible with a bias as small as possible. However, in his conclusion, Cragg (1983) says that “the major problem encountered was downward bias of the covariance matrices”. It is clear that results for the ERP and power of tests and for the bias of the variance estimator concern fundamentally different objects (see Flachaire 2002). In this section, we study the finite sample bias of the Cragg estimators of the covariance matrix, as the number of instruments increases. We will see that this problem is closely related to the choice of the number of instruments. Our study will give some evidence which could be helpful for using the Cragg estimators in practice.

A condition of identification requires that the number of instruments must be greater than or equal to the number of regressors and less than or equal to the number of observations: $k \leq l \leq n$. Thus, when the regressors are taken as instruments with no additional instruments, $l = k$, the Cragg estimator (10) reduces to the Eicker-White estimator (6). In the opposite case, when there are as many instruments as regressors, $l = n$, the Cragg estimator (10) reduces to

$$\widehat{\text{Var}}(\hat{\beta}_n) = (\mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{X})^{-1} \quad (29)$$

However, this last estimator is inconsistent, because $\hat{\Omega}$ is not a consistent estimator of Ω . In order to study the bias of the Cragg estimator, we compare realizations of the Cragg variance estimator (10) with the asymptotic variance, defined by $\text{Var}(\hat{\beta}_{\text{GMM}}) =$

$(\mathbf{X}^\top \mathbf{W}(\mathbf{W}^\top \boldsymbol{\Omega} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X})^{-1}$, using the following statistic

$$\mathcal{B} = \frac{\widehat{\text{Var}}(\hat{\beta}_1) - \text{Var}(\hat{\beta}_1)}{\text{Var}(\hat{\beta}_1)} \quad (30)$$

where $\widehat{\text{Var}}(\hat{\beta}_1)$ is the Cragg estimator of the variance of $\hat{\beta}_1$, see (15), and $\text{Var}(\hat{\beta}_1)$ is the asymptotic variance of $\hat{\beta}_1$, that is, a diagonal element of $\text{Var}(\hat{\beta}_{\text{GMM}})$. In our experiments, we consider different choices of instruments. Let us assume that \mathbf{X}_d , \mathbf{X}_{cd} and \mathbf{X}_{cp} are respectively the matrix of the inverses, cross-divisions and cross-products of the regressors. Thus, we define the following matrix of instruments as:

$$\begin{aligned} l = k : & \quad \mathbf{W}_0 = \mathbf{X} \\ l = k + 5 : & \quad \mathbf{W}_1 = (\mathbf{X}, \mathbf{X}_d, \mathbf{X}_{cd}) \\ l = k + 22 : & \quad \mathbf{W}_2 = (\mathbf{W}_1, \mathbf{X}_{cp}, \mathbf{X}^2, \mathbf{X}_d^2, \mathbf{X}_{cd}^2, \mathbf{X}_{cp}^2, \mathbf{X}^3, \mathbf{X}_d^3, \mathbf{X}_{cd}^3, \mathbf{X}_{cp}^3) \\ l = k + 38 : & \quad \mathbf{W}_3 = (\mathbf{W}_2, \mathbf{X}^4, \mathbf{X}_d^4, \mathbf{X}_{cd}^4, \mathbf{X}_{cp}^4, \mathbf{X}^5, \mathbf{X}_d^5, \mathbf{X}_{cd}^5, \mathbf{X}_{cp}^5) \end{aligned}$$

In our model design (14), we have $\mathbf{X}_d = (1/\mathbf{x}_1, 1/\mathbf{x}_2)$, $\mathbf{X}_{cd} = (\mathbf{x}_1/\mathbf{x}_2, \mathbf{x}_2/\mathbf{x}_1, 1/(\mathbf{x}_1\mathbf{x}_2))$ and $\mathbf{X}_{cp} = \mathbf{x}_1\mathbf{x}_2$. Thus, \mathbf{W}_0 , \mathbf{W}_1 , \mathbf{W}_2 and \mathbf{W}_3 have respectively 0, 5, 22 and 38 additional instruments. The case $l = n$ is defined in (29) and corresponds to 97 additional instruments.

Figure 8 shows the Empirical Distribution Functions (EDFs) of 10,000 realizations of \mathcal{B} for the Cragg estimators based on the HC_3 transformation of the restricted residuals. The bias would be given by the first moment of the EDF, but more information can be drawn from the EDF to measure the quality of the covariance matrix estimator, as for instance the dispersion: a distribution centered around zero with a small dispersion would give realizations of Cragg's covariance estimator always quite close to the asymptotic variance estimator. We can see from this figure that with no additional instruments, that is, for the Eicker-White test, the distribution is highly skewed with a large dispersion (line $l = k$). The addition of a few instruments shows a symmetry around zero with a small dispersion (line $l = k + 5$). Furthermore, we can see that adding too many instruments moves the distribution to the left and generates asymmetry. In other words, our results suggest that the bias of the Cragg estimator of the variance decreases with a few additional instruments and its precision increases. However, this estimator becomes biased to the left when too many additional instruments are used.

6 Conclusion

The Eicker-White heteroskedasticity consistent covariance matrix estimator, for the Ordinary Least Squares parameter estimator, is commonly used in practice. Our experimental results suggest that tests based on this estimator can perform very badly in finite samples: the ERP can be large and a lack of power can be very considerable compared to the efficient GLS test. Cragg (1983) proposed a more efficient estimator, which exhibits large errors in the rejection probability in finite samples, and thus cannot be used in practice.

Our simulation results suggest that if we compute the Cragg estimator with a transformation of the restricted residuals and with the wild bootstrap, and if we use inverses/cross-products/cross-divisions/power of regressors as instruments, we can obtain more reliable and more efficient tests. In addition, our results suggest that the use of a few additional instruments can lead to very good finite sample performance, but the use of too many instruments introduces a significant bias. Finally, our results make clear that more reliable and more efficient inferences can be obtained in the presence of heteroskedasticity of unknown form, with an appropriate use of the Cragg estimator and with a few additional instruments.

Our study has been focused on testing hypotheses. A key result is to use restricted residuals rather than unrestricted residuals. However, it is usually convenient to compute confidence intervals, for which no restricted regression model is specified. Confidence intervals and hypothesis tests are closely related and, in order to benefit from the refinement of restricted residuals, we should construct confidence intervals based on inverting hypothesis tests, as proposed by Davidson (2000) and Davidson and MacKinnon (2001).

We have limited our study to the case of cross-section data sets, with exogenous regressors and unconditional heteroskedasticity. Cumby, Huizinga, and Obstfeld (1983) extended the Cragg estimators to include the case of endogenous regressors and Gonçalves and Kilian (2004) proposed an extended wild bootstrap method that is valid in the presence of conditional heteroskedasticity of unknown form. Further investigations should be made in this more generalized case to extend these results in time series models.

References

- Beran, R. (1986). Discussion of “Jackknife bootstrap and other resampling methods in regression analysis” by C.F.J. Wu. *Annals of Statistics* 14, 1295–1298.
- Chesher, A. and I. Jewitt (1987). “The bias of a heteroskedasticity consistent covariance matrix estimator”. *Econometrica* 55, 1217–1222.
- Cragg, J. G. (1983). “More efficient estimation in the presence of heteroskedasticity of unknown form”. *Econometrica* 51, 751–63.
- Cumby, R. E., J. Huizinga, and M. Obstfeld (1983). “Two-step two-stage least squares estimation in models with rational expectations”. *Journal of Econometrics* 21, 333–55.
- Davidson, R. (2000). “Bootstrap confidence intervals based on inverting hypothesis tests”. *Econometric Reviews* 19, 49–54.
- Davidson, R. and E. Flachaire (2001). “The wild bootstrap, tamed at last”. working paper IER#1000, Queen’s University.
- Davidson, R. and J. G. MacKinnon (1985). “Heteroskedasticity-robust tests in regression directions”. *Annales de l’INSEE* 59/60, 183–218.
- Davidson, R. and J. G. MacKinnon (2000). “Bootstrap tests: How many bootstraps?”. *Econometric Reviews* 19, 55–68.

- Davidson, R. and J. G. MacKinnon (2001). "Improving the reliability of bootstrap confidence intervals". paper presented at the conference *Resampling Methods in Econometrics* held at the Université de Montréal in October.
- Eicker, B. (1963). "Limit theorems for regression with unequal and dependant errors". *Annals of Mathematical Statistics* 34, 447–456.
- Flachaire, E. (1999). "A better way to bootstrap pairs". *Economics Letters* 64, 257–262.
- Flachaire, E. (2002). "Bootstrapping heteroskedasticity consistent covariance matrix estimator". *Computational Statistics* 17, 501–506.
- Flachaire, E. (2004). "Propriétés en échantillon fini des tests robustes à l'hétéroscédasticité de forme inconnue". *Annales d'Économie et Statistique*, forthcoming.
- Flachaire, E. (2005). "Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap". *Computational Statistics and Data Analysis*, forthcoming.
- Freedman, D. A. (1981). "Bootstrapping regression models". *Annals of Statistics* 9, 1218–1228.
- Godfrey, L. G. and C. D. Orme (2001). "Significance levels of heteroskedasticity-robust tests for specification and misspecification: some results on the use of wild bootstraps". paper presented at ESEM'2001, Lausanne.
- Godfrey, L. G. and C. D. Orme (2004). "Controlling the finite sample significance levels of heteroskedasticity-robust tests of several linear restrictions on regression coefficients". *Economics Letters* 82, 281–287.
- Gonçalves, S. and L. Kilian (2004). "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form". *Journal of Econometrics* 123, 89–120.
- Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators". *Econometrica* 50, 1029–54.
- Hansen, L. P. (1985). "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators". *Journal of Econometrics* 30, 203–38.
- Horowitz, J. L. (1997). "Bootstrap methods in econometrics: theory and numerical performance". In *Advances in Economics and Econometrics: Theory and Application*, Volume 3, pp. 188–222. David M. Kreps and Kenneth F. Wallis (eds), Cambridge, Cambridge University Press.
- Horowitz, J. L. (2000). "The bootstrap". In *Handbook of Econometrics*, Volume 5. J. J. Heckman and E. E. Leamer (eds), Elsevier Science.
- Liu, R. Y. (1988). "Bootstrap procedure under some non-i.i.d. models". *Annals of Statistics* 16, 1696–1708.
- Long, J. S. and L. H. Ervin (2000). "Heteroscedasticity consistent standard errors in the linear regression model". *The American Statistician* 54, 217–224.
- MacKinnon, J. G. (2002). "Bootstrap inference in econometrics". *Canadian Journal of Economics* 35, 615–645.
- MacKinnon, J. G. and H. L. White (1985). "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties". *Journal of Econometrics* 21, 53–70.

- van Giersbergen, N. P. A. and J. F. Kiviet (2002). “How to implement bootstrap hypothesis testing in static and dynamic regression model: test statistic versus confidence interval approach”. *Journal of Econometrics* 108, 133–56.
- White, H. (1980). “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity”. *Econometrica* 48, 817–838.
- Wu, C. F. J. (1986). “Jackknife bootstrap and other resampling methods in regression analysis”. *Annals of Statistics* 14, 1261–1295.

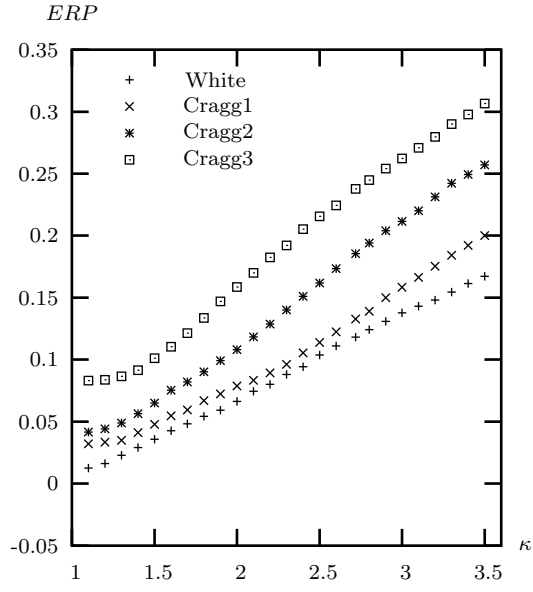


Fig 1: Asymptotic-unrestricted residuals

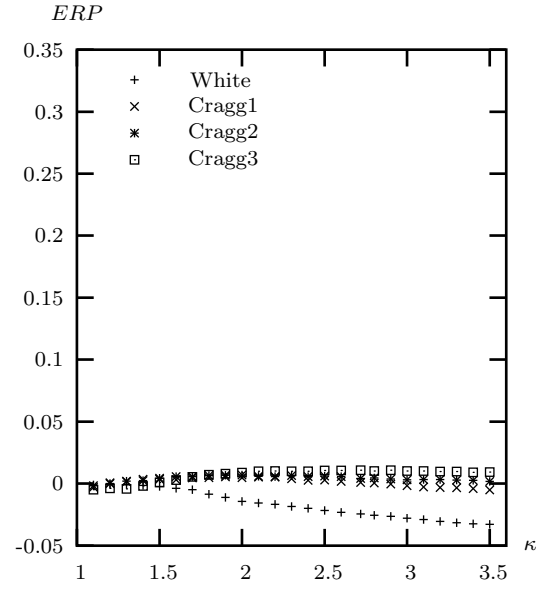


Fig 2: Asymptotic-restricted residuals

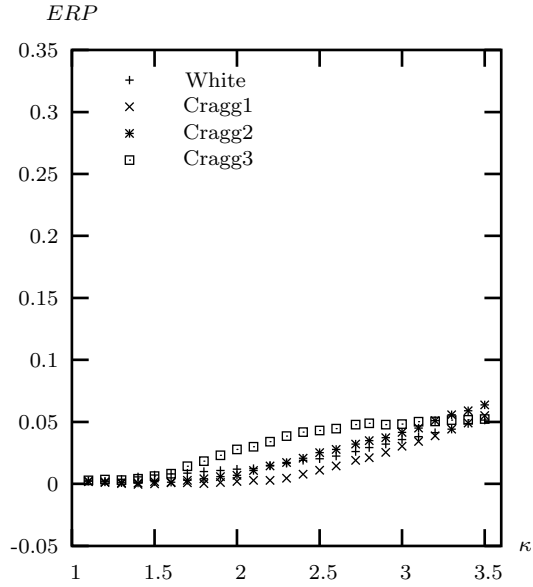


Fig 3: Bootstrap-unrestricted residuals

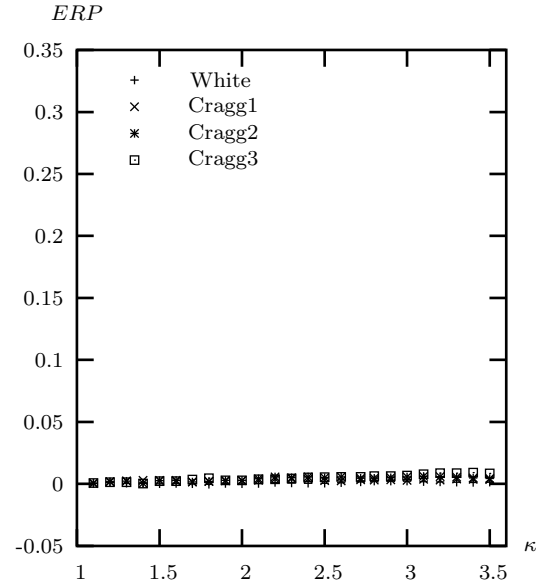


Fig 4: Bootstrap-restricted residuals

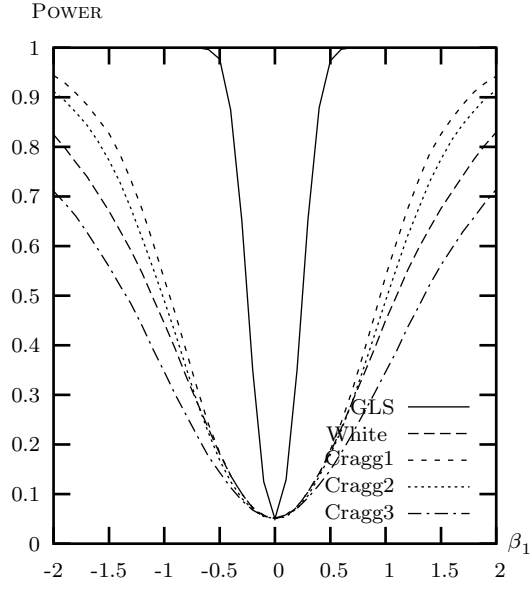


Fig 5: Bootstrap-unrestricted residuals

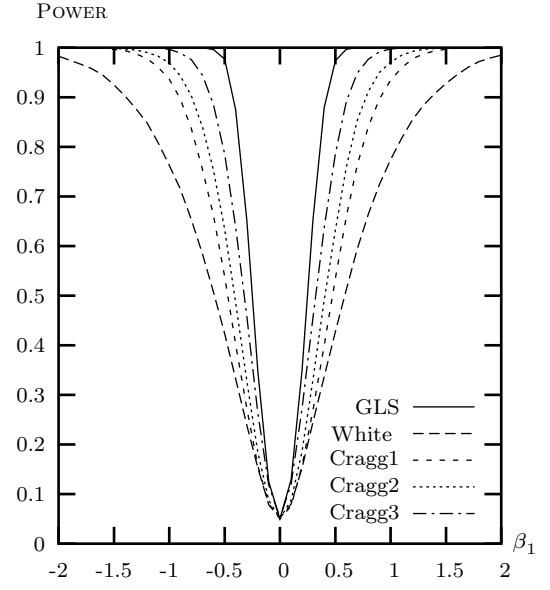


Fig 6: Bootstrap-restricted residuals

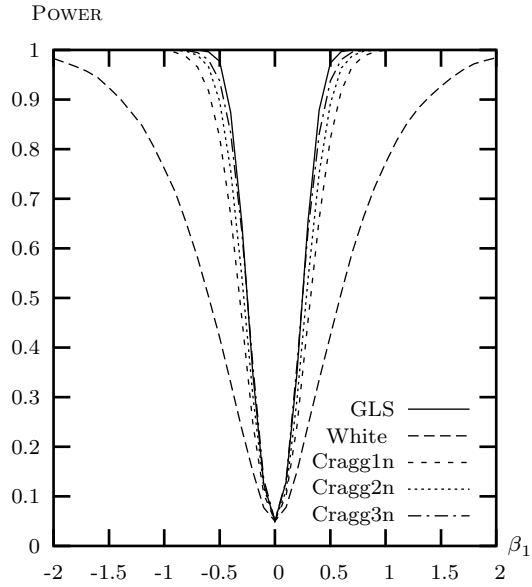


Fig 7: New instruments

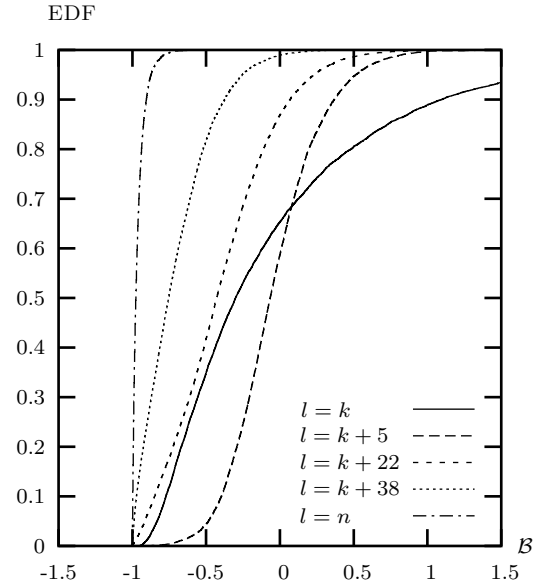


Fig 8: Bias study: EDFs of \mathcal{B}